Homework Assignment # 2

Due: Monday, February 25, 2019, 11:59 p.m.
Total Points: 100

This assignment is designed to help you better understand concepts that were presented during class. You may work in small groups (2 or 3 individuals) on the homework assignments, but it is expected that you submit your own work. Therefore, **each student is responsible for submitting their own homework**, whether they work in groups or individually. The submitted homework must include the name of other individuals you worked with (if applicable), as well as all resources that were used to solve the problem (e.g. web sites, books, research papers, other people, etc.). Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

Your assignment must be submitted as a single pdf document on Canvas. The questions and your answers must be typed; for example, in Latex, Microsoft Word, Lyx, etc. Images may be scanned and inserted into the document if it is too complicated to draw them properly. Be sure to show all work.

All code (if applicable) should be turned in when you submit your assignment. Use Matlab, Python, R, Java or C. Be sure to include an complete explanation on how to run/execute your code.

All assignments must be submitted on time to receive credit. No late work will be accepted, unless you have a prior arrangement with the instructor.

Question 1. [15 POINTS]

Let $X_1, \ldots, X_n$ be i.i.d. random variables with probability distribution

$$f(x|\alpha) = \alpha^x (1-\alpha)^{1-x}$$

where $x \in \{0, 1\}$. Assuming that the unknown parameter α was selected from a $(0, 1)$-uniform distribution, find the Bayes estimator of α.

Question 2. [10 POINTS]

Let X have an exponential density

$$f(x|\theta) = \begin{cases} \theta e^{-\theta x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Suppose that n samples $x_1, \ldots, x_n$ are drawn independently according to $f(x|\theta)$. What is the maximum-likelihood estimate for θ?

Question 3. [15 POINTS]

Suppose X is observed data whose observation depends on two hypothesis (e.g. H_0 and H_1). Depending on whether H_0 or H_1 is true, X is defined as follows:

$$\begin{aligned} &\text{if } H_0 \text{ is true} : X = aN - s \\ &\text{if } H_1 \text{ is true} : X = bN + s \end{aligned} \tag{2}$$

where, $s > 0$ $a > 0$, and $b > 0$ are fixed real numbers and N is a continuous random variable with density

$$f_N(n) = \frac{1}{\pi(1+n^2)}, \quad n \in \mathbb{R} \tag{3}$$

(a) Assuming uniform cost and equal priors, find the Bayesian decision rule for classifying between H_0 and H_1.

(b) What is the minimum Bayesian risk?

Question 4. [10 POINTS]

Let Y be an exponentially distributed random variable with parameter α. This problem is about estimating Y, based on the observation X, which is another exponential random variable, whose parameter is identical to the realization $Y = y$. Namely,

$$f_{X|Y}(x|y) = \begin{cases} ye^{-yx}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

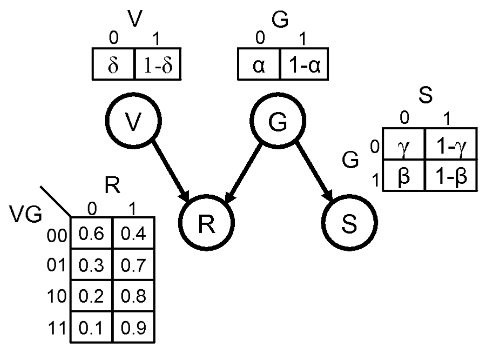
(a) Find $f_{Y|X}(y|x)$. Hint: Perform integration by parts to simplify

(b) Find the MAP estimate, $\hat{Y}_{\text{MAP}}$ of Y.

(c) Find the MMSE estimate, $\hat{Y}_{\text{MMSE}}$ of Y, given $X = x$. Note that if Z is an exponentially distributed random variable with parameter λ, then its nth moment is, $E[Z^n] = \frac{n!}{\lambda^n}$.

Question 5. [10 POINTS]

In this problem, there are 4 binary random variables: G = "gray", V = "Vancouver", R="rain", and S="sad". Consider the directed graphical model, shown below, which describes the relationship between these random variables:



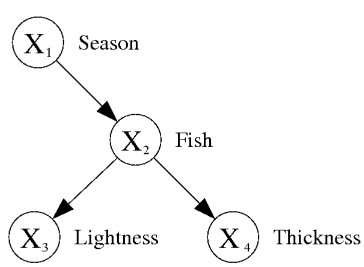
(a) Write an expression for $P(S = 1|V = 1)$ in terms of $\alpha, \beta, \gamma, \delta$.

(b) Write an expression for $P(S = 1|V = 0)$. Is this the same or different to $P(S = 1|V = 1)$? Explain why.

Question 6. [15 POINTS]

Consider the Bayesian network shown below. Here, the nodes represent the following variables

$$\begin{aligned} X_1 &\in \{winter, spring, summer, autumn\} \\ X_2 &\in \{salmon, seabass\} \\ X_3 &\in \{light, medium, dark\} \\ X_4 &\in \{wide, thin\} \end{aligned} \tag{5}$$



The corresponding conditional probability tables are:

$$P(X_1) = \begin{bmatrix} .25 & .25 & .25 & .25 \end{bmatrix} \quad P(X_2|X_1) = \begin{bmatrix} .9 & .1 \\ .3 & .7 \\ .4 & .6 \\ .8 & .2 \end{bmatrix}$$
$$P(X_3|X_2) = \begin{bmatrix} .33 & .33 & .34 \\ .8 & .1 & .1 \end{bmatrix} \quad P(X_4|X_2) = \begin{bmatrix} .4 & .6 \\ .95 & .05 \end{bmatrix} \tag{6}$$

Note that in $P(X_4|X_2)$, the rows represent X_2 and the columns represent X_4 (so each row sums to one and represents the child of the CPD). Thus, $P(X_4 = thin|X_2 = \text{sea bass}) = 0.05, P(X_4 = thin|X_2 = salmon) = 0.6$, etc.

(a) Suppose the fish was caught on December 20th - the end of autumn and the beginning of winter - and thus let $P(X_1) = \begin{bmatrix}.5 & 0 & 0 & .5\end{bmatrix}$, instead of the above prior. (This is called soft evidence, since we do not know the exact value of X_1, but we have a distribution over it.) Suppose the lightness has not been measured but it is known that the fish is wide. Classify the fish as salmon or sea bass. Assume uniform costs. Clearly show all steps.

(b) Suppose all we know is that this fish is wide and dark lightness. What season is it now, most likely? Use $P(X_1) = \begin{bmatrix}.25 & .25 & .25 & .25\end{bmatrix}$

Question 7. [10 POINTS]

Moments of a mixture of Gaussians. Consider a mixture of K Gaussians, where $\mathbf{x}$ describes a multi-dimensional random variable, where π_k is the prior probability of class k (e.g. $\pi_k \equiv P(c_k)$)

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{7}$$

(a) Show that $E[\mathbf{x}] = \sum_k \pi_k \boldsymbol{\mu}_k$

(b) Show that $cov[\mathbf{x}] = \sum_k \pi_k \left(\boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T\right) - E[\mathbf{x}]E[\mathbf{x}]^T$. Hint: use the fact that $cov[\mathbf{x}] = E[\mathbf{x}\mathbf{x}^T] - E[\mathbf{x}]E[\mathbf{x}]^T$. $cov[\mathbf{x}]$ represents the covariance of $\mathbf{x}$. Using our notation from class, this is the same as $Cov(\mathbf{X}, \mathbf{X})$

Question 8. [15 POINTS]

This exercise will give you some practice coding the EM algorithm. We will apply the mixture of Gaussian technique to one-dimensional data . You will find a data set on Canvas: data1(txt). Create a program to estimate the means, standard deviations, and weights of a mixture of Gaussians via the EM algorithm. You will probably want to have three functions: one that performs the expectation step, one that performs the maximization step, and one that runs the outer loop. Be sure to include all code and plots in your homework submission. **You must only use your own code to solve this problem. Do not download code or use pre-defined libraries or packages.**

The log-likelihood for a mixtures of Gaussians is as follows:

$$L = \sum_{j=1}^{N} \log\left(p(x_j)\right) = \sum_{j=1}^{N} \log\left(\sum_{k=1}^{K} P(c_k)\mathcal{N}(x_j|\mu_k, \sigma_k)\right) \tag{8}$$

(a) Data set 1 was created with 3 Gaussians. Run your EM algorithm. Give the resulting means, standard deviations, and weights (that is, the model parameters). Plot the log likelihood of the data as it changes during training (i.e. plot log-likelihood against the number of iterations of EM so far). Use a stopping threshold of 0.001.

(b) How does the log likelihood change if you use 2 or 4 Gaussians? What happens to the parameters of the model?